Transcriptional regulation constrains the organization of genes on eukaryotic chromosomes

Sarath Chandra Janga*†, Julio Collado-Vides‡, and M. Madan Babu*†

*Laboratory of Molecular Biology, Medical Research Council, Hills Road, Cambridge CB2 0QH, United Kingdom; and ‡Programa de Genomica Computacional, Centro de Ciencias Genomicas, Universidad Nacional Autonoma de Mexico, Apartado Postal 565-A, Av Universidad, Cuernavaca, Morelos, 62100 Mexico D.F., Mexico

Edited by Aaron Klug, Medical Research Council, Cambridge, United Kingdom, and approved August 21, 2008 (received for review July 1, 2008)

Genetic material in eukaryotes is tightly packaged in a hierarchical manner into multiple linear chromosomes within the nucleus. Although it is known that eukaryotic transcriptional regulation is complex and requires an intricate coordination of several molecular events both in space and time, whether the complexity of this process constrains genome organization is still unknown. Here, we present evidence for the existence of a higher-order organization of genes across and within chromosomes that is constrained by transcriptional regulation. In particular, we reveal that the target genes (TGs) of transcription factors (TFs) for the yeast, *Saccharomyces cerevisiae*, are encoded in a highly ordered manner both across and within the 16 chromosomes. We show that (*i*) the TGs of a majority of TFs show a strong preference to be encoded on specific chromosomes, (*ii*) the TGs of a significant number of TFs display a strong preference (or avoidance) to be encoded in regions containing particular chromosomal landmarks such as telomeres and centromeres, and (*iii*) the TGs of most TFs are positionally clustered within a chromosome. Our results demonstrate that specific organization of genes that allowed for efficient control of transcription within the nuclear space has been selected during evolution. We anticipate that uncovering such higher-order organization of genes in other eukaryotes will provide insights into nuclear architecture, and will have implications in genetic engineering experiments, gene therapy, and understanding disease conditions that involve chromosomal aberrations.

gene order | genome | nuclear architecture | systems biology | network

Although transcription in both prokaryotes and eukaryotes involves the evolutionarily conserved core RNA polymerase subunit, the whole process of transcriptional regulation is fundamentally different. In contrast to prokaryotes where transcription primarily relies on the *cis*-regulatory DNA sequences alone (1), eukaryotic transcription is regulated at least at three major levels (2, 3). The first is at the level of DNA sequence where the transcription factor (TF) associates with *cis*-regulatory elements to regulate transcription of the relevant gene (2). The second is at the level of chromatin, which allows segments within a chromosomal arm to switch between different transcriptional states, that is, between a state that suppresses transcription and one that allows for gene activation (2). This involves changes in nucleosome occupancy and chromatin structure, both of which are controlled by the interplay between remodeling complexes, histone modification, DNA methylation, and a variety of repressive and activating mechanisms (4, 5). The third is at the level of nuclear architecture, which includes organization of chromosomes into chromosomal territories, and the dynamic, temporal, and spatial organization of specific chromosomal loci–all of which are known to influence gene expression (6–11). Thus, unlike in prokaryotes, transcription in eukaryotes is an energy-intensive, multistep process, involving a large number of molecular events to be coordinated both in space and time [supporting information (SI) Fig. S1]. Given the intricacy involved in a single transcriptional regulatory interaction, one can ask whether or not the complexity of the whole network of transcriptional interactions (Fig. S1) has imposed a significant constraint on the organization of genes across the different eukaryotic chromosomes. This becomes particularly interesting in the light of a recent work that demonstrated that tuning the expression level of a single gene could provide an enormous fitness advantage to an individual in a population of cells (12). Thus, one could extrapolate that optimization of transcriptional regulation on a global scale, such as the efficient expression of relevant genes under specific conditions, would have significant advantage on the fitness of an individual in a genetically heterogeneous population.

Although several studies have reported that genes with similar expression patterns cluster on the genome and that gene order is conserved, no study has investigated whether transcriptional regulation has constrained organization of genes across and within the chromosomes; in particular, whether the set of genes regulated by a given TF are (*i*) randomly distributed across different chromosomes or encoded on specific chromosomes, (*ii*) distributed in an unbiased manner within a chromosomal arm or display preference to be encoded in regions containing particular chromosomal landmarks, or (*iii*) positionally clustered within a chromosome. Here, we investigate these questions by using the recently available genome-scale data on 13,853 high-confidence regulatory interactions (Fig. S1). These data cover 156 TFs and 4,495 target genes for the model eukaryote *Saccharomyces cerevisiae*, whose genetic material is organized into 16 linear chromosomes.

Results and Discussion

The Majority of TFs Show a Strong Preference to Regulate Genes on Specific Chromosomes. Several elegant studies have elucidated that the organization of chromosomes within the eukaryotic nucleus is nonrandom and that they occupy distinct volumes called chromosomal territories (6, 7). In yeast, in addition to the ordered movements during cell division, it has been demonstrated that interphase chromosomes undergo large rapid movements (>0.5 μm in a 10-s interval; nuclear diameter of ≈ 2 μm) and that such movements could reflect the metabolic state of the cell (Fig. S1; refs. 7 and 9 and references therein). These observations have suggested that the nonrandom organization of the chromosomes could (*i*) allow functional compartmentalization of the nuclear space, thus potentially enhancing or repressing expression of specific genes, and (*ii*) bring coregulated genes into physical proximity to coordinate gene expression. The above-mentioned observations on the nonrandom nuclear architecture and chromosomal dynamics together with the fact that transcriptional regulation in eukaryotes is an energy-

Author contributions: S.C.J. and M.M.B. designed research; S.C.J. and M.M.B. performed research; J.C.-V. contributed new reagents/analytic tools; S.C.J. and M.M.B. analyzed data; and S.C.J. and M.M.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

†To whom correspondence may be addressed. E-mail: sarath@mrc-lmb.cam.ac.uk or madanm@mrc-lmb.cam.ac.uk

This article contains supporting information online at www.pnas.org/cgi/content/full/0806317105/DCSupplemental.

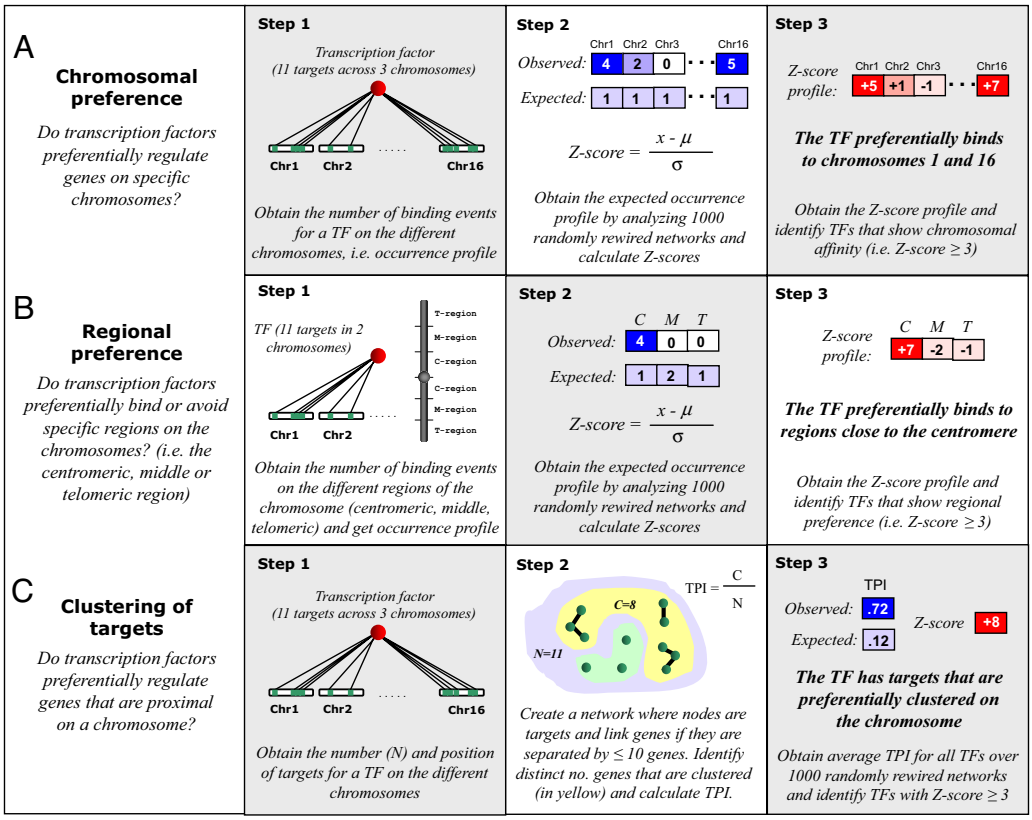

Fig. 1. Schematic illustration showing the methods used to estimate the significance for chromosomal preference (*A*), regional preference (*B*), and clustering of target genes (*C*). See *Materials and Methods* for details. *x*, observed value; μ, mean; σ, standard deviation.

intensive, highly coordinated, and time-intensive process motivated us to ask whether such considerations have constrained the positioning of genes in specific chromosomes during the course of evolution.

We first investigated whether the targets of TFs tend to be preferentially encoded on specific chromosomes. We therefore analyzed the chromosomal location of the targets for each TF in the currently available map of protein–DNA interactions for yeast (Fig. 1*A* and *Materials and Methods*). We first created a "chromosome preference profile" for every TF, which is a vector that contains the number of target genes on each of the 16 chromosomes. By comparing this vector with what is expected by chance (see *Materials and Methods*), we identified the TFs that displayed a significant preference to have their targets on specific chromosomes more often than what is expected by chance.

Because the null model is critical to obtain statistical significance, we ensured that the random networks are as close as possible to the real network in terms of the topology and the gene distribution on the chromosomes. The random networks were therefore obtained by employing a rewiring procedure, preserving the connectivity distribution and the inherent chromosomal distribution of the genes. In other words, the number of targets for each TF and the number of TFs regulating a given target gene (TG) in the random networks will be the same as what is seen in the real network, but the interactions between them are randomly rewired. Because this procedure does not randomize the chromosomal position of a gene, any inherent, nonrandom clustering of genes on the genome is explicitly maintained. Furthermore, this procedure treats every chromosome independently by maintaining the same gene density and the same number of genes as seen in the real yeast chromosomes. This procedure therefore allows us to assess any preference for binding by the TFs analyzed in our study. For all observations reported here, statistical significance was assessed based on *P* value and *Z* score. Only TFs with $P \leq 10^{-3}$ and $|Z| \geq 3$ were considered to show a significant difference compared with the null model. To correct for multiple testing, we calculated *q* values as a measure of significance by using the *q* value package in R. We estimate a false discovery rate (FDR) of 0.3% when calling all $P \leq 10^{-3}$ as significant (see Table S1).

Through this analysis, we found that a majority of the TFs (84 TFs, $P < 10^{-3}$ and $Z \geq 3$) showed a striking preference to encode a significant fraction of their TGs on at least one particular chromosome. Of these, 78% (66 TFs) showed preference to only one chromosome, 18% (15 TFs) showed preference to two chromosomes, and a smaller fraction (4%) of the TFs showed preference to three or more chromosomes. Fig. 2 shows all of the 16 chromosomes of *S. cerevisiae* along with the TFs that have been identified to preferentially bind to the target loci on them. Our investigation identified several TFs that have a strong preference to regulate genes on specific chromosomes (Table S1). Some of these include (*i*) the global regulatory hub Sok2p, showing a significant preference for binding to chromosome XV (observed: *x*, 67; expected: μ, 32; Z, 6.7; $P < 10^{-3}$) and regulating genes important for pseudohyphal differentiation and vesicle trafficking; (*ii*) Phd1p, showing a preference for binding to chromosome V (*x*, 52; μ, 23; Z, 7.0; $P < 10^{-3}$) and chromosome IX (*x*, 32; μ, 14; Z, 5.1; $P < 10^{-3}$) and controlling expression of genes required for differentiation; and (*iii*) Msn4p, showing preference for chromosome XIII (*x*, 32; μ, 13; Z, 5.4; $P < 10^{-3}$), regulating expression of genes involved in stress response. Although it is interesting to note that all of the 16 chromosomes have a preferred set of TFs binding them (Fig. 2*B*), the number of TFs showing preference to a particular chromosome does not correlate with the physical size of the chromosome (in bp), gene content, or the gene density (Table S2). Taken together, these observations indicate that the targets of most TFs are not randomly distributed across the different chromosomes. Instead, they are highly ordered and show a preference to be encoded on specific chromosomes, independent of the size and the gene density of the chromosome.

Our finding that such a pattern of organization exists for the

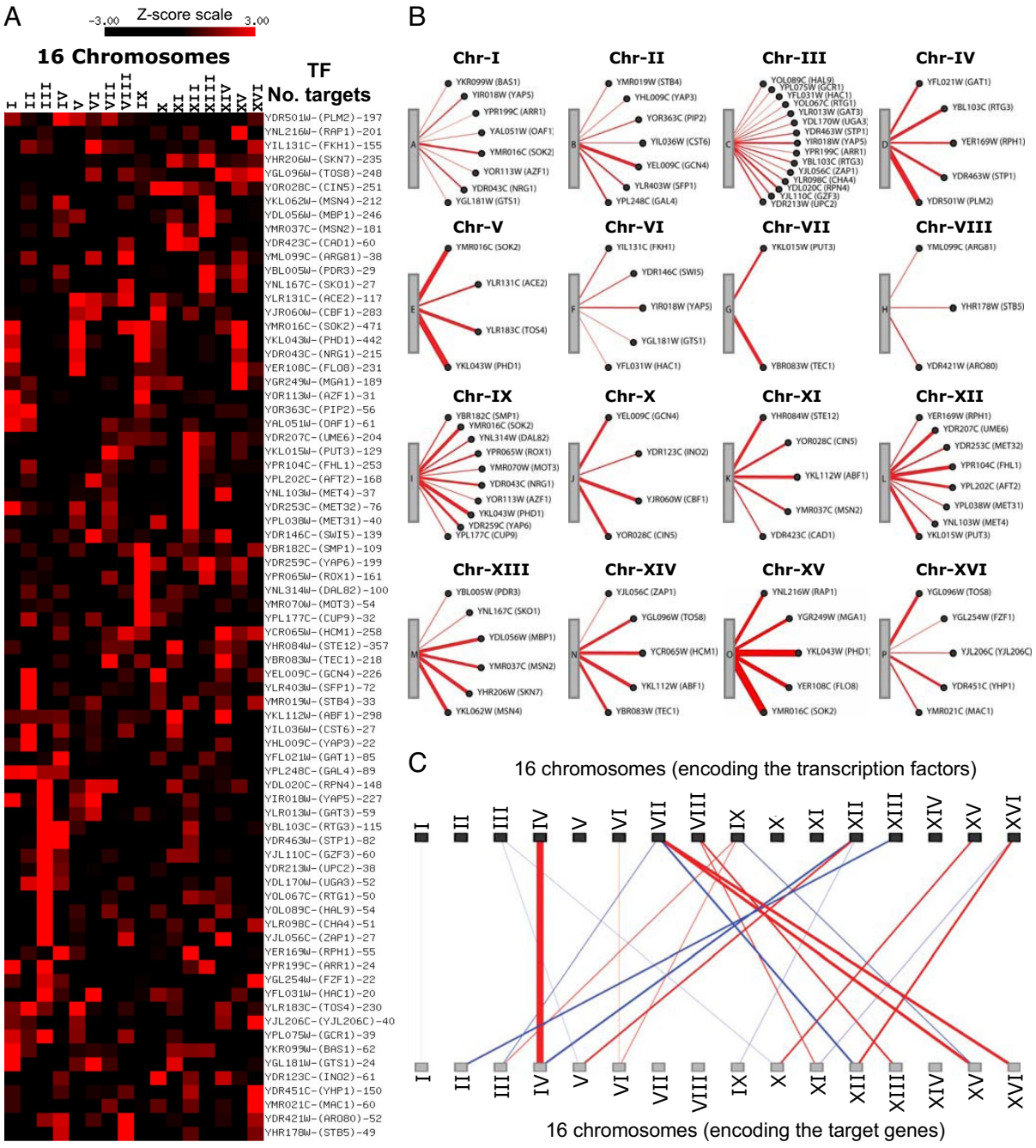

Fig. 2. Chromosomal preference for binding by TFs. (*A*) Each column in the matrix represents one of the 16 chromosomes labeled I to XVI. Each row represents the *Z* score significance profile of a particular TF (shown on the right) to have its targets on the different chromosomes (see Fig. 1*A*). The top 75 TFs (selected by *P* value and higher *Z* scores) are ordered after hierarchically clustering their *Z* score profiles. The number of target genes is mentioned next to the gene name. (*B*) TFs with target preference for each of the 16 chromosomes. Only those TFs that show preference for binding to chromosomes with *Z* scores ≥ 3, $P \leq 10^{-3}$, and regulate >16 genes are shown. Each chromosome has a set of TFs that tend to preferentially bind them. The thickness of the red line is proportional to the absolute number of target genes for that TF on the chromosome. (*C*) Higher-order organization of regulatory interactions. The top column denotes the chromosomes where the TFs are encoded, and the bottom column denotes the chromosomes where the target genes are encoded. Red and blue lines connecting the two chromosomes mean that TFs originating from a specific chromosome tend to preferentially encode or avoid targets on a particular chromosome, respectively. The thickness is proportional to the *Z* score.

distribution of targets of TFs motivated us to analyze (*i*) whether the TFs themselves show a preference to be encoded on specific chromosomes, and in particular, whether global regulatory proteins show any such preference, and (*ii*) whether there are any patterns of higher-order organization of regulatory interactions between chromosomes. Our investigation on the first question unambiguously revealed that TFs and particularly the global regulatory hubs do not show any preference to be encoded on specific chromosomes. Instead, the distribution was similar to what is expected by chance (Table S2). However, we identified the existence of a higher-order organization of regulatory interactions wherein several TFs that are encoded on specific chromosomes tend to preferentially regulate or avoid regulating genes on distinct chromosomes. Fig. 2*C* shows the links between chromosomes that display a statistically significant tendency to either interact (red line; $P < 10^{-3}$; $Z \geq 3$) or avoid interaction (blue line; $P < 10^{-3}$; $Z \leq -3$)

in the context of transcriptional regulation. These observations suggest that TFs encoded in specific chromosomes can show distinct preferences to regulate targets encoded on particular chromosomes and might reflect a coordinated, and possibly a combinatorial, effect between TFs that are encoded in the same chromosome.

A Significant Number of the TFs Tend to Have Targets on Specific Regions of the Chromosomal Arm. Apart from the fact that the nucleus is organized into subcompartments, creating microenvironments that facilitate distinct nuclear functions, several studies that visualized precise chromosomal loci have revealed that specific regions of the chromosomes display restricted displacement to varying degrees (Fig. S1; refs. 7 and 9). For instance, in yeast, chromosomal "landmarks" such as the telomeres and centromeres show marked constraints in their movements within the nuclear space when compared with other chromosomal loci. In addition, live microscopy studies have revealed that centromeres tend to cluster near the spindle pole body (SPB), whereas the telomeres tend to be tethered to the nuclear envelope (7, 9). Moreover, it has been shown that yeast chromosomes form chromosomal loops, where the telomeric ends come closer to each other than to the centromeres. Such anchoring of chromosomal regions is thought to be reversible and is known to involve microtubules that associate with the SPB (for centromeres) and the yKu heterodimeric protein, Esc1p and Sir4p (for telomeres) (7, 9). This phenomenon of periodic attachment of distinct regions of the chromosomal arms to the nuclear periphery appears to be a conserved mechanism and is believed to regulate patterned gene expression, possibly by separating transcriptionally active and inactive chromosomal domains (13, 14). These observations motivated us to assess whether such phenomena, during the course of evolution, could have constrained the target genes of TFs to be encoded within distinct regions of the chromosomal arm.

In particular, we asked whether TFs tend to preferentially bind or avoid specific regions on the linear chromosomes, such as regions closer to the centromere or the telomere, or the regions in-between. To investigate this question, we first divided each chromosomal arm into three equal regions (in bp): C, containing the centromere; M, the middle region; and T, containing the telomere. For each TF, we then created a "regional preference profile," which contains the number of targets in each of the three regions. Comparing these results with random expectation allowed us to assess the statistical significance (see Fig. 1*B* and *Materials and Methods*). This enabled the discovery of TFs that display a significant bias to bind to particular regions of the chromosomal arm independent of the specific chromosome. We found that 29 TFs (Fig. 3*A*) showed a statistically significant preference ($P < 10^{-3}$; $Z \geq 3$ at a FDR of 0.5%) to bind to a particular region over others, thus providing the first evidence for the prevalence for such an effect (see Table S3). We show that several TFs display a strong preference to bind specific regions on chromosomal arms. For instance, Hsf1p, the trimeric heat shock regulatory protein, and Msn2p, the multicopy suppressor of SNF1 mutation protein, tend to preferentially regulate genes that are encoded in regions closer to the centromere, whereas the bZIP domain containing TFs Yap5p and Yap6p, which are required under stress conditions, tend to bind to regions closer to the telomere. Additional evidence came from our inspection of the TFs that avoided binding to a particular region (Fig. 3*B*). We found that certain TFs like the osmosis-dependent regulators Skn7p and Msn2p clearly avoided binding to the T region, whereas the pleiotropic drug regulators Pdr1p and Smp1p avoided regulating genes in the C region. Interestingly, the suppressor of kinase Sok2p, which regulates genes involved in cellular differentiation, avoids binding to both the C and M regions of the chromosomes, displaying a clear preference to bind to the region containing the telomere. Taken together, these observations suggest that events that allowed clustering of certain functionally related genes, based on their usage, accessibility and transcriptional activity, have been

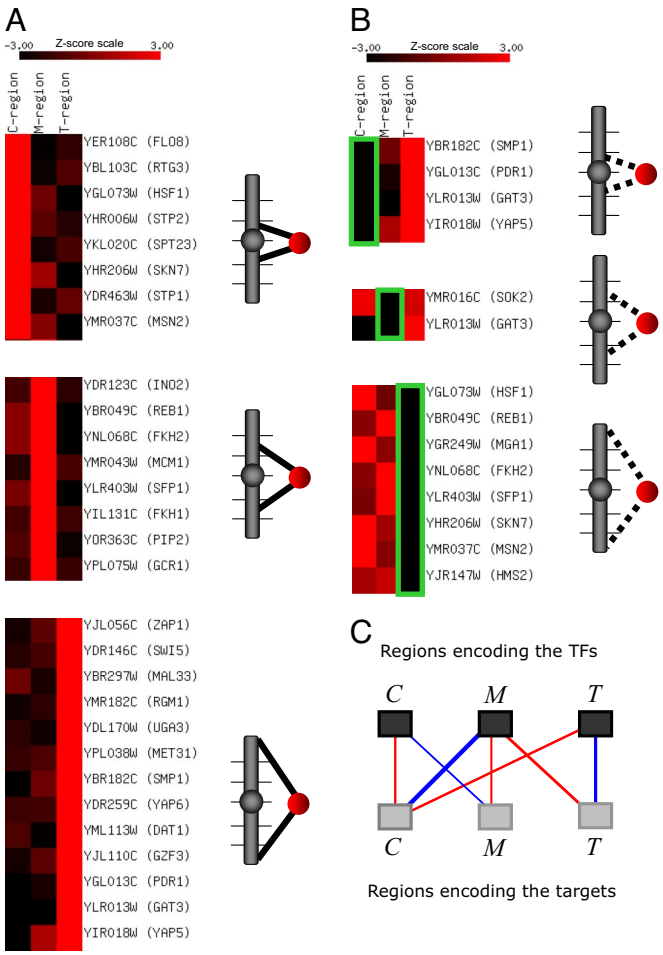

Fig. 3. TFs showing significant regional preference or avoidance for binding on the chromosomes (see Fig. 1*B*). (*A*) TFs that show a strong tendency to have their targets on the C region, M region, or T region on the chromosome. (*B*) TFs that show a strong avoidance to have their targets on the three regions. Green boxes highlight the group of TFs that show significant regional avoidance for one of the three regions. In the diagram next to the matrices, thick black lines indicate preference and broken black lines indicate avoidance. Only TFs with $P < 10^{-3}$ and $|Z| \geq 3$ are shown in both cases. (*C*) Higher-order organization of regulatory interactions. The top column denotes regions on the chromosomal arm where the TFs are encoded and the bottom column denotes the regions where the targets are encoded. Lines connecting the two regions mean that TFs originating from a specific region tend to preferentially have (red lines) or avoid (blue lines) targets on a particular region of the chromosome. The thickness is proportional to the *Z* score.

selected during evolution. Consistent with this proposal, it is interesting to note that regions that cluster at the nuclear periphery such as the telomeres, as well as the mating-type loci, are generally transcriptionally silent, whereas internally located regions encoding metabolic enzymes on the chromosomal arm get recruited to nuclear pores on transcriptional activation (15–18).

We then investigated whether (*i*) the loci encoding TFs and, in particular, global regulatory proteins, show any regional preference and (*ii*) there are patterns of higher-order organization of regulatory interactions involving specific chromosomal regions, that is, whether TFs encoded in specific regions tend to preferentially regulate genes on other chromosomal regions. Although our investigation along these lines revealed the absence of any such preferential organizational pattern for the loci encoding TFs (see Table S4), we discovered that genes encoding global regulatory hubs tend to strongly avoid being encoded in regions closer to the telomere ($P = 0.004$). Investigations to uncover the presence of

higher-order interactions between specific chromosomal regions revealed that TFs encoded elsewhere in the genome regulate genes within the T region, whereas TFs within the T region appear to preferentially avoid regulating genes in the same region ($P = 0.007$; Fig. 3*C*). These observations are consistent with the fact that genes on telomeric and subtelomeric regions are largely repressed. Because of this phenomenon, rearrangement events that resulted in the TFs and hubs (which regulate genes in the T region) to be encoded elsewhere on the chromosome would have been favored during evolution. Given the dynamic nature of the different chromosomal regions and the differential transcriptional activity associated with specific regions, such organization of loci encoding TFs within specific chromosomal regions, and patterns of higher-order regulatory interactions may have been selected during evolution. Taken together, the findings reported here suggest that such regional preferences are not only seen for the targets of specific TFs, but also for global regulatory hubs and the regulatory interactions affecting expression of genes in specific chromosomal regions.

Most TFs Show a Strong Preference to Positionally Cluster Their Targets Within a Chromosome. Next, we investigated whether the target genes of a TF are proximal to each other on the chromosome or whether they are relatively far apart within the same region. Although several studies have revealed that genes with similar expression profiles (coexpressed genes) cluster on the chromosome (19–21), very few studies have addressed whether the targets of the same TF cluster on the chromosome on a genomic scale (22). Although previous studies have unambiguously revealed the existence of chromosomal domains that contain genes with similar expression pattern (coexpressed genes), it should be kept in mind that clustering of coexpressed genes need not always imply regulation by the same TF, because coexpressed genes may be clustered for several reasons, such as mechanisms involving chromatin remodeling, transcriptional readthrough, regulation of genes by the same TF, or regulation by different TFs in the same transcriptionally active euchromatic domain (23). Therefore, we initiated a systematic investigation and analyzed whether the targets of TFs display positional clustering on a given chromosome or not.

We first defined and calculated the Target Proximity Index (TPI) for each TF (see *Materials and Methods* and Fig. 1*C*). In short, the TPI for a TF represents the fraction of all of the regulated genes that show proximal clustering on the chromosome. In our study we defined proximity, D, as the number of genes that separate two targets of a TF. We then compared the TPI values for the observed and the random networks to obtain the statistical significance. From our analysis, we found that most TFs (>75%) showed high TPI values (TPI >0.6, $P < 10^{-3}$; at a FDR of 0.1% for D ≤ 20), suggesting a strong preference for target genes to be clustered within a distance range of ≈20 genes. On the contrary, TPI values in random networks for the same distance threshold were found to be significantly <0.2. To ensure that the observations are not biased by (*i*) gene duplication events that might result in nearby duplicate genes being regulated by the same TF or (*ii*) divergent, bidirectional genes that might artificially increase the TPI score, we carried out the relevant control calculations. In the filtered network, we removed (*i*) all tandem duplicates from our dataset and (*ii*) randomly chose a target gene from a divergent, bidirectional gene pair and calculated the TPI score. Our results did not change after controlling for tandem duplicates and bidirectionally transcribed genes (see Table S5 and Fig. S2) suggesting that what we observe are truly attributable to positional clustering of targets on a chromosome. An investigation of how many genes are positionally clustered within the window of 20 genes revealed that, on an average, such a window only contains 2.6 genes regulated by the same TF. This is striking and suggests that all three mechanisms, that is, (*i*) chromatin remodeling, (*ii*) regulation by different TFs in the same euchromatic domain, and (*iii*) regulation by the same TF within a euchromatic domain, may contribute to the previously observed domains of coexpressed genes.

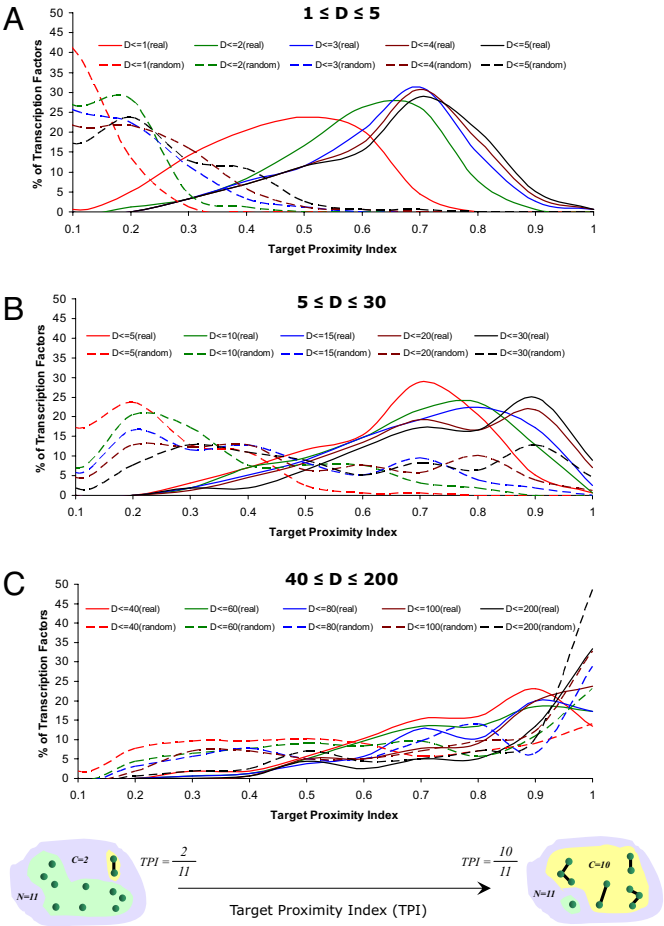

Fig. 4. Frequency distribution of TPI values. Distribution of Target Proximity Index (TPI) for all TFs in the real and randomly constructed networks at different proximity values, that is, D values (see *Materials and Methods*) are shown: (*A*) D ≤ 1 to D ≤ 5; (*B*) D ≤ 5 to D ≤ 30; and (*C*) D ≤ 40 to D ≤ 200. Note that, in the real network, the maximum proportion of TFs have TPI values that are much higher than what is seen for the random networks (at ≈0.8 for real network and 0.2 for random networks at D ≤ 20), demonstrating that most TFs show clustering of their targets in this distance range.

To validate the robustness of our definition of proximity on the TPI values, we systematically varied this parameter (D) from 1 to 200 and compared them against what was obtained in random networks (Fig. 4). We found that significant separation between real data and random networks occurred for the definition of proximity (D) as being <20 genes, suggesting that this could reflect the average size of a possible open euchromatic domain that is available for transcription in yeast. Our results therefore suggest that evolution might have favored certain recombination events that allowed genes that need to be regulated by the same TF to be encoded close to each other. Another distinct possibility given that transcriptional regulatory networks are likely to be plastic (24) would be that selection could have first driven clustering of genes that need to be coregulated and then new transcriptional regulatory interactions could have evolved afterward. Regardless of the driving force, the evolutionary advantages are clear: such a clustering of targets would not demand high concentrations of TFs in the nucleus that are generally expressed in low quantities and prevent inappropriate regulation of unrelated target genes. Such an organization has the added advantage of minimizing noise in expression

levels, which has been recently proposed to be an additional driving force for gene order conservation (25).

Conclusions

In conclusion, our study demonstrates that the complexity of transcriptional regulation constrains genome organization at several levels. Specifically, our findings illustrate that (*i*) a majority of the TFs tend to preferentially encode their targets on only one or a few chromosomes, (*ii*) several TFs display a strong preference for regulating genes on specific regions on the chromosomal arms, and (*iii*) most TFs tend to bind targets that are positionally clustered within a specific region on the chromosome. Our findings that TFs encoded in specific chromosomes and within distinct regions show a strong preference to regulate genes on distinct chromosomes (and regions) open up several questions and expand our need to understand eukaryotic gene regulation at a higher level. The findings reported here are consistent with several molecular mechanisms, such as the genome-wide loop model of chromosomes (26), the presence of expression hubs (27) and transcription factories (28, 29), and the nuclear gating hypothesis (30).

With the development of experimental methods such as chromosome conformation capture (3C), 4C, 5C, and 6C, and the availability of genome-scale data on protein–DNA interactions from high-throughput experiments in other eukaryotes, our work provides a fundamental framework by which such questions can be systematically studied for higher eukaryotes. In fact, a preliminary analysis in mammalian systems by using stem cell differentiation factors Sox2, Oct4, and Nanog has indeed revealed a striking preference for these TFs to encode their targets on specific chromosomes (see *SI Text* and Fig. S3). We therefore believe that our work, which demonstrates that gene organization is constrained by the process of transcriptional regulation in yeast, is likely to be a paradigm that is also applicable to other eukaryotes.

The findings reported here have several direct applications. For instance, the map that we describe for yeast in this study can serve as a guide and can be exploited in genetic engineering experiments for identifying the most appropriate region (on the 16 chromosomes) to incorporate a gene of interest, in particular, if it has to be regulated under the control of a specific TF. Describing such maps for higher eukaryotes will have implications in gene therapy and in rationally identifying suitable sites to incorporate a gene of interest to produce transgenic organisms. We anticipate that revealing the presence of such patterns of organization of genes within the linear chromosomes of eukaryotes, such as humans, will have significant implications in our understanding of transcriptional regulation, chromosomal territories, their role in cellular differentiation, and of specific chromosomal disorders, such as recombination events and copy number variations that are prevalent in diverse diseases such as cancer.

Materials and Methods

Transcriptional regulatory network for *S. cerevisiae* was assembled from the results of literature curation of ChIP-chip experiments (see ref. 31 and references therein). This network consists of 156 TFs and 4,495 TGs comprising of 13,853 interactions. The top 20% of the TFs with highest outgoing connectivity were defined as hubs. Chromosomal positions of genes were obtained from http://www.yeastgenome.org. Tandem duplicates and bidirectionally transcribed genes were identified using a blast E-value cutoff of 10^{-2} and chromosomal coordinates of genes. Chromosomal, regional, and proximity preferences of TFs were calculated as shown in Fig. 1 and are explained in *SI Materials and Methods*. Statistical significance of the observations was estimated by comparison against 1,000 random networks (see *SI Materials and Methods* for details).

ACKNOWLEDGMENTS. We thank A. Wuster, R. Pache, E. Levy, S. De, P. Kota, M. Vadivelu, V. Pisupati, L. Lo Conte, K. Ganesh, G. Moreno-Hagelsieb, A. Pombo, S. Teichmann, K. Nagai, K. J. Patel, L. Hurst, J. Chin, J. Kilmartin, P. Dear, C. Chothia, R. Henderson, A. Travers, H. McMahon, S. Munro, V. Ramakrishnan, and A. Klug for providing helpful comments. We apologize to colleagues whose relevant work could not be cited because of lack of space. This work was supported by Medical Research Council-Laboratory of Molecular Biology (M.M.B. and S.C.J.) and by Cambridge Commonwealth Trust (S.C.J.).

1. Browning DF, Busby SJ (2004) The regulation of bacterial transcription initiation. *Nat Rev Microbiol* 2(1):57–65.
2. Lee TI, Young RA (2000) Transcription of eukaryotic protein-coding genes. *Annu Rev Genet* 34:77–137.
3. van Driel R, Fransz PF, Verschure PJ (2003) The eukaryotic genome: A system regulated at different hierarchical levels. *J Cell Sci* 116(Pt 20):4067–4075.
4. Millar CB, Grunstein M (2006) Genome-wide patterns of histone modifications in yeast. *Nat Rev Mol Cell Biol* 7(9):657–666.
5. Lieb JD, Clarke ND (2005) Control of transcription through intragenic patterns of nucleosome composition. *Cell* 123(7):1187–1190.
6. Cremer T, Cremer C (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2(4):292–301.
7. Gasser SM (2002) Visualizing chromatin dynamics in interphase nuclei. *Science* 296(5572):1412–1416.
8. Misteli T (2004) Spatial positioning: A new dimension in genome function. *Cell* 119(2):153–156.
9. Akhtar A, Gasser SM (2007) The nuclear envelope and transcriptional control. *Nat Rev Genet* 8(7):507–517.
10. Fraser P, Bickmore W (2007) Nuclear organization of the genome and the potential for gene regulation. *Nature* 447(7143):413–417.
11. Kosak ST, *et al.* (2007) Coordinate gene regulation during hematopoiesis is related to genomic organization. *PLoS Biol* 5(11):e309.
12. Dekel E, Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. *Nature* 436(7050):588–592.
13. Finlan LE, *et al.* (2008) Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet* 4(3):e1000039.
14. Guelen L, *et al.* (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature.*
15. Taddei A, *et al.* (2006) Nuclear pore association confers optimal expression levels for an inducible yeast gene. *Nature* 441(7094):774–778.
16. Cabal GG, *et al.* (2006) SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature* 441(7094):770–773.
17. Casolari JM, *et al.* (2004) Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. *Cell* 117(4):427–439.
18. Ishii K, Arib G, Lin C, Van Houwe G, Laemmli UK (2002) Chromatin boundaries in budding yeast: The nuclear pore connection. *Cell* 109(5):551–562.
19. Hurst LD, Pal C, Lercher MJ (2004) The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* 5(4):299–310.
20. Cohen BA, Mitra RD, Hughes JD, Church GM (2000) A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet* 26(2):183–186.
21. Spellman PT, Rubin GM (2002) Evidence for large domains of similarly expressed genes in the Drosophila genome. *J Biol* 1(1):5.
22. Hershberg R, Yeger-Lotem E, Margalit H (2005) Chromosomal organization is shaped by the transcription regulatory network. *Trends Genet* 21(3):138–142.
23. Batada NN, Urrutia AO, Hurst LD (2007) Chromatin remodelling is a major source of coexpression of linked genes in yeast. *Trends Genet* 23(10):480–484.
24. Borneman AR, *et al.* (2007) Divergence of transcription factor binding sites across related yeast species. *Science* 317(5839):815–819.
25. Batada NN, Hurst LD (2007) Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat Genet* 39(8):945–949.
26. Francastel C, Schubeler D, Martin DI, Groudine M (2000) Nuclear compartmentalization and gene activity. *Nat Rev Mol Cell Biol* 1(2):137–143.
27. Kosak ST, Groudine M (2004) Gene order and dynamic domains. *Science* 306(5696):644–647.
28. Cook PR (1999) The organization of replication and transcription. *Science* 284(5421):1790–1795.
29. Osborne CS, *et al.* (2004) Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* 36(10):1065–1071.
30. Blobel G (1985) Gene gating: A hypothesis. *Proc Natl Acad Sci USA* 82(24):8527–8529.
31. Balaji S, Babu MM, Iyer LM, Luscombe NM, Aravind L (2006) Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *J Mol Biol* 360(1):213–227.